

Computer-Aided Malignancy Risk Assessment of Nodules in Thyroid US Images Utilizing Boundary Descriptors

Michalis A. Savelonas¹, Dimitris E. Maroulis¹, Dimitris K. Iakovidis², and Nikos Dimitropoulos³

^{1,2}*Dept. of Informatics and Telecommunication, University of Athens, Greece*

³*Dept. of Medical Imaging, Euromedica Medical Center, Athens, Greece*

¹*rtsimage@di.uoa.gr*

²*dimitris.iakovidis@ieee.org*

Abstract

This paper investigates the diagnostic potential of boundary descriptors calculated on delineations of thyroid nodules in ultrasound images. The utilized set of boundary descriptors includes compactness, chain code histogram and fractal dimension. Experiments were conducted on thyroid US images, so as to evaluate the discriminating capability of each boundary descriptor for the classification of thyroid nodules in terms of their malignancy risk. The receiver operating characteristic (ROC) is used for the evaluation of the experimental results. It is derived that thyroid nodules can be classified according to their malignancy risk, by utilizing their compactness, fractal dimension and chain code histogram, obtaining an area under curve (AUC) which reaches 0.93.

1. Introduction

The advent of high-resolution ultrasound (US) imaging as a fast, non-invasive, low-cost diagnostic tool has facilitated the acquisition of detailed information associated with physiological or pathological findings of the thyroid gland [1],[2]. Nodule size, echogenicity, as well as the presence of irregular or ill-defined boundaries have been proved to correlate with thyroid malignancy risk [3],[4]. Boundary descriptors calculated on nodule delineations, could be integrated within a computer-aided diagnosis tool, contributing to the objectification of medical decisions.

Chain Coding (CC), introduced by Freeman [5], is a common boundary representation approach. Many applications using CC representation have been reported up to date [6]-[8], however, to the best of our knowledge, CC or CC derivatives have not been considered for the definition of boundary descriptors of US findings. This could be attributed to the consideration that since every small deviation along the

boundary is encoded, CC is expected to be extremely sensitive to US speckle noise. However, the existence of accurate US segmentation methods [1],[9] demonstrating noise robustness encourages the application of CC derived boundary descriptors, such as CC histogram itself and compactness, on US findings.

Compactness (CMP) is a typical measure of the space-filling nature of a boundary. A high CMP value is associated with an irregular boundary, since it indicates a large perimeter enclosing a small area. Typically, benign masses are expected to have lower values of CMP as compared to malignant tumors [10]. Kallergi et al [11] suggest CMP as a discriminative boundary descriptor for assessing malignancy risk of breast nodules. For the same medical problem, Horsch et al [12] considered normalized radial gradient. However, this morphologic feature involves image gradients, which are sensitive to US speckle noise.

Fractal dimension (FD) indicates the presence of self-similarity at several levels of magnification and constitutes another measure of boundary irregularity. Microlobulated or highly spiculated contours of malignant tumors are expected to demonstrate fractal behavior, which should be absent in the case of benign masses [13],[14]. Rangayyan and Nguyen [13] computed FD from the one-dimensional signature of a given mass, defined as the Euclidean distance of each contour point to the centroid of the contour. FD is a well-defined mathematical entity specified in terms of limiting processes. However, its computation using real, discrete data, such as images, is not obvious. Much of the literature regarding fractal analysis has been concerned with the estimation of FD given a discrete data set [14].

The focus of this paper is to investigate the diagnostic potential of CC histogram, CMP and FD, so as to facilitate the malignancy risk assessment of thyroid nodules in US images. The classification experiments are evaluated by means of the receiver

operating characteristic (ROC). To the best of our knowledge, a computer-aided diagnosis method for thyroid malignancy risk assessment based on boundary descriptors has not yet been proposed.

The rest of this paper is organized in three sections. Section II briefly reviews CC histogram, CMP and FD. The results from their application for malignancy risk assessment of thyroid nodules in US images are apposed in Section III, and the conclusions of this study are summarized in Section IV.

2. Boundary descriptors

2.1. Descriptors derived from chain codes

The core idea of CC involves the movement along a digital curve or along a sequence of border pixels based on n -connectivities. The direction of each movement is encoded by means of a numbering scheme ($\{i|i=0, 1, \dots, n\}$ denoting an angle of $360 \cdot i/n$), clockwise or counter-clockwise, from the positive x -axis, as shown in Fig.1. Thus, CC can be viewed as a connected sequence of straight-line segments with specified lengths and directions.

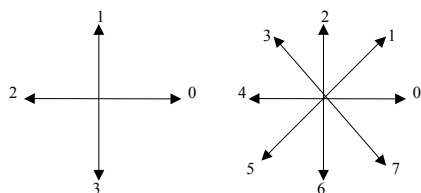


Figure 1. Directions of a 4-directional CC and of an 8-directional CC.

The CC can be normalized with respect to the starting point by a straightforward procedure: it is treated as a circular sequence of direction numbers and the starting point is redefined so that the resulting sequence of numbers forms an integer of minimum magnitude. Normalization with respect to rotation can be obtained by using the first difference of the CC instead of the code itself. This difference is obtained by counting the number of direction changes that separate two adjacent elements of the code. For instance, the first difference of the 4-direction chain code 10103322 is 3133030. The code can be treated as a circular sequence by computing the first element of the difference using the transition between the last and first components of the chain. Accordingly, the result in this example is 33133030 [15]. The histogram describing the frequency of occurrence of each CC element could

be used as a boundary descriptor. Moreover, since a CC is a complete representation of a curve, it can be used to calculate boundary descriptors of an object, such as perimeter, area and CMP.

CMP is a dimensionless quantity providing a measure of contour complexity, whereas it is independent of translation, rotation, and scale. For this study, the following definition of CMP is adopted:

$$CMP = \frac{P^2}{A} \quad (1)$$

where P is the perimeter and A is the area of the object, calculated with the use of CC. According to Eq. (1), CMP has a minimum value of 4π and increases for elongated shapes or shapes with rough contours.

2.2. Fractal dimension

The FD can be computed utilizing the box counting algorithm [16]. This algorithm is based on partitioning the shape image into square boxes of size $L \times L$ and counting the number of boxes ($n(L)$) containing a portion of the shape. The FD is calculated as the absolute value of the slope of the line obtained from the linear regression of the $(\log(L), \log(n(L)))$ curve, by varying the box size L (Fig. 2).

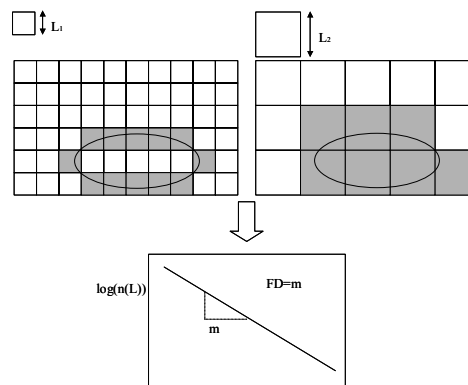


Figure 2. Illustration of the box counting method.

3. Experimental results

Thyroid US examinations were performed on 312 patients using a digital US imaging system HDI 3000 ATL with a 5-12 MHz linear transducer in the radiology department of Euromedica diagnostic center in Greece. A total of 173 longitudinal in vivo digital images of hypoechoic thyroid nodule cases were acquired at a resolution of 256×256 pixels with a 256 gray-level depth.

A special purpose software suite in Microsoft Visual C++ was developed and executed on a 3.2 GHz Intel Pentium IV workstation for the extraction of the CC histogram, CMP and FD of thyroid nodules on ultrasound images.

The study is limited to cases of hypoechoic nodules, since hypoechogenicity in conjunction with boundary irregularity demonstrate an extremely high correlation with malignancy risk [17]. Three expert physicians characterized the boundary of each hypoechoic nodule used in the experiments as either regular or irregular, following the classification of Tomimori [3]. The dataset used in the experiments comprised of 131 medium-risk hypoechoic nodules characterized by regular boundaries and 42 high-risk hypoechoic nodules characterized by irregular boundaries.

The Variable Background Active Contour model [9] is applied for the delineation of the thyroid nodules. The model constants used in the experiments were chosen as: $\lambda^+ = 5$, $\lambda^- = 5$, $\mu = 650$ and $a = 10^{-13}$. Figure 3 illustrates example delineations obtained for nodules of regular (Fig. 3(a)-3(c)), as well as of irregular (Fig. 3(d)-3(e)) boundaries.

Experiments were performed for the classification of thyroid nodules with respect to their malignancy risk. The utilized set of boundary descriptors includes CC histogram, CMP and FD, calculated on the delineation of each nodule. A k -nearest neighbor (k -NN) classifier [18] was used for the classification. The value k was experimentally determined as $k=5$.

Table 1 illustrates the area under curve (AUC) derived from the ROC of each combination of the utilized boundary descriptors. The highest AUC is obtained by the combined use of CC histogram, CMP and FD, indicating that each of these features provides complementary boundary information. It could be noticed that FD obtains the lowest AUC, when compared to the AUC obtained by each individually applied descriptor. This could be attributed to the relatively low image resolution, which limits the variety of levels of magnification utilized for the calculation of FD. However, the inclusion of FD in the utilized boundary descriptor combination results in an increased AUC. The ROC curve associated with the combination of all boundary descriptors is illustrated in Fig. 4.

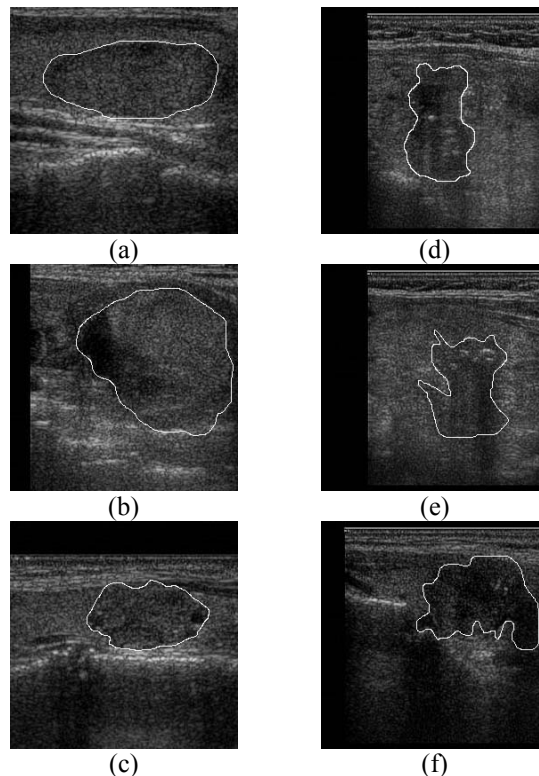


Figure 3. Thyroid US images with delineated nodules. (a-c) nodules of regular boundaries, (d-f) nodules of irregular boundaries.

Table 1. Area under curve for each descriptor combination

Feature Vector	AUC
{CMP}	0.82
{FD}	0.76
{CC histogram}	0.79
{CMP, FD}	0.84
{CMP, CC histogram}	0.86
{FD, CC histogram}	0.81
{CMP, FD, CC histogram}	0.93

4. Conclusions

In this paper, we investigated the potential of boundary descriptors for the assessment of thyroid nodules on US images according to malignancy risk. We performed classification experiments utilizing the k -NN classifier for feature vectors derived from all combinations of CC histogram, CMP and FD.

The experimental evaluation on real thyroid US images lead to the following conclusions: 1) boundary descriptors of thyroid nodules are capable of discriminating between medium-risk and high-risk nodules, with an AUC that reaches 0.93, and 2) CC

histogram, CMP and FD, all provide complementary boundary information of thyroid nodules in US images.

Future perspectives of this work include the investigation of features encoding local intensity distribution so as to facilitate the classification of nodules characterized by blurred or ill-defined boundaries.

Acknowledgments. We would like to thank EUROMEDICA S.A., Greece, for the provision of the thyroid ultrasound images and their contribution in the evaluation of the results. This work was supported by the Greek General Secretariat of Research and Technology and the European Social Fund, through the PENED 2003 program (grant no. 03-ED-662).

5. References

- [1] S. Tsantis, N. Dimitropoulos, D. Cavouras, and G. Nikiiforidis, "A Hybrid Multi-Scale Model for Thyroid Nodule Boundary Detection on Ultrasound Images", *Comp. Meth. Progr. Biom.*, vol. 64, 2006, pp. 86-98.
- [2] P. Gimondo, P. Mirk, G. Messina, G. Pizzi, A. Tomei, "The Role of Ultrasonography in Thyroid Disease", *Minerva Med.*, vol. 84, 1993, pp. 671-680.
- [3] E.K. Tomimori, R.Y. Camargo, H. Bisi, G. Medeiros-Neto, "Combined Ultrasonographic and Cytological Studies in the Diagnosis of Thyroid Nodules", *Biochimie*, vol. 81, no. 5, 1999, pp. 447-452.
- [4] E. Koike, S. Noguchi, H. Yamashita, T. Murakami, A. Ohshima, H. Kawamoto, H. Yamashita, "Ultrasonographic Characteristics of Thyroid Nodules: Prediction of Malignancy", *Arch. Surg.*, vol. 136, 2001, pp. 334-337.
- [5] H. Freeman, "On the Encoding of Arbitrary Geometric Configurations", *IRE Trans. Electron. Comput.*, vol. 10, 1961, pp. 260-268.
- [6] Y.K. Liu, B. Zalik, "An Efficient Chain Code with Huffman Coding", *Patt. Rec.*, vol. 38, 2005, pp. 553-557.
- [7] P. Zingaretti, M. Gasparroni, L. Vecci, "Fast Chain Coding of Region Boundaries", *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 20, no. 4, 1998, pp. 407-414.
- [8] F.Y. Shih, W.T. Wong, "A One-Pass Algorithm for Local Symmetry of Contours from Chain Code", *Patt. Rec.*, vol. 32, no. 7, 1999, pp. 1203-1210.
- [9] D.E. Maroulis, M.A. Savelonas, D.K. Iakovidis, S.A. Karkanis, N. Dimitropoulos, "Variable Background Active Contour Model for Computer-Aided Delineation of Nodules in Thyroid Ultrasound Images", *IEEE Trans. Inf. Tech. Biom.*, vol. 11, no. 5, 2007, pp. 537-543.
- [10] K. Nakamura et al, "Computerized Analysis of the Likelihood of Malignancy in Solitary Pulmonary Nodules with use of Artificial Neural Networks", *Radiol.*, vol. 14, no. 3, 2000, pp. 823-830.
- [11] M. Kallergi, "Computer-Aided Diagnosis of Mammographic Microcalcification Clusters", *Med. Phys.*, vol. 31, no.2, 2004, pp. 314-326.
- [12] K. Horsch, M.L. Giger, L.A. Venta, C.J. Vyborny, "Automatic Segmentation of Breast Lesions on Ultrasound", *Med. Phys.*, vol. 28, no. 8, 2001, pp. 1652-1659.
- [13] R.M. Rangayyan, T.M. Nguyen, "Fractal Analysis of Contours of Breast Masses in Mammograms", *J. Dig. Im.*, vol. 20, no. 3, 2007, pp. 223-237.
- [14] Y. Tao, E.C.M. Lam, Y. Tang, "Feature Extraction Using Wavelet and Fractal", *Patt. Rec. Lett.*, vol. 22, no. 2001, pp. 271-287.
- [15] R.C. Gonzales, R.E. Woods, *Digital image processing*, 3rd edition, Prentice Hall, 2008.
- [16] H.O. Peitgen and D. Saupe, *The Science of Fractal Images*, Springer Verlag, 1998.
- [17] E. Papini, R. Guglielmi, A. Bianchini, A. Crescenzi, S. Taccogna, F.Nardi, C. Panunzi, R. Rinaldi, V. Toscano, and C. M. Pacella, "Risk of Malignancy in Nonpalpable Thyroid Nodules: Predictive Value of Ultrasound and Color-Doppler Features", *J. Clinic. Endocr. Metabol.*, vol. 87, no. 5, 2002, pp. 1941-1946.
- [18] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 3rd edition, Academic Press, 2006.